

Regularity Model for Noisy Multiobjective Optimization

Wang, Handing; Zhang, Qingfu; Jiao, Licheng; Yao, Xin

DOI:

[10.1109/TCYB.2015.2459137](https://doi.org/10.1109/TCYB.2015.2459137)

License:

Creative Commons: Attribution (CC BY)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Wang, H, Zhang, Q, Jiao, L & Yao, X 2015, 'Regularity Model for Noisy Multiobjective Optimization', *IEEE Transactions on Cybernetics*, no. 99. <https://doi.org/10.1109/TCYB.2015.2459137>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Regularity Model for Noisy Multiobjective Optimization

Handing Wang, *Student Member, IEEE*, Qingfu Zhang, *Senior Member, IEEE*,
Licheng Jiao, *Senior Member, IEEE*, and Xin Yao, *Fellow, IEEE*

Abstract—Regularity models have been used in dealing with noise-free multiobjective optimization problems. This paper studies the behavior of a regularity model in noisy environments and argues that it is very suitable for noisy multiobjective optimization. We propose to embed the regularity model in an existing multiobjective evolutionary algorithm for tackling noises. The proposed algorithm works well in terms of both convergence and diversity. In our experimental studies, we have compared several state-of-the-art of algorithms with our proposed algorithm on benchmark problems with different levels of noises. The experimental results showed the effectiveness of the regularity model on noisy problems, but a degenerated performance on some noisy-free problems.

Index Terms—Local principal component analysis (PCA), multiobjective optimization, noise, regularity model.

I. INTRODUCTION

A MULTIOBJECTIVE optimization problem (MOP) can be stated as follows:

$$\begin{aligned} \min \quad & F(x) = (f_1(x), \dots, f_m(x)) \\ \text{s.t.} \quad & x \in \Omega \end{aligned} \quad (1)$$

Manuscript received January 24, 2015; revised May 20, 2015; accepted July 16, 2015. This work was supported in part by the National Basic Research Program (973 Program) of China under Grant 2013CB329402, in part by EPSRC under Grant EP/J017515/1 on “DAASE: Dynamic Adaptive Automated Software Engineering,” in part by the Program for Cheung Kong Scholars and Innovative Research Team in University under Grant IRT11170, in part by the National Natural Science Foundation of China under Grant 61329302, Grant 91438103, Grant 61473241, and Grant 91438201, and in part by the Fund for Foreign Scholars in University Research and Teaching Programs (the 111 Project) under Grant B07048. The work of X. Yao was supported by the Royal Society Wolfson Research Merit Award. This paper was recommended by Associate Editor H. Ishibuchi.

H. Wang was with Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, International Research Center of Intelligent Perception and Computation, Xidian University, Xi'an 710071, China. She is now with the Department of Computing, University of Surrey, Guildford GU2 7XH, U.K. (e-mail: wanghanding.patch@gmail.com).

Q. Zhang is with the Department of Computer Science, City University of Hong Kong, Hong Kong, and also with the School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, U.K. (e-mail: qzhang@essex.ac.uk).

L. Jiao is with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an 710071, China (e-mail: lchjiao@mail.xidian.edu.cn).

X. Yao is with the School of Computer Science, University of Birmingham, Birmingham B15 2TT, U.K. (e-mail: x.yao@cs.bham.ac.uk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCYB.2015.2459137

where Ω is the decision space and F is the objective function consisting of m individual objectives f_1, \dots, f_m . Let $x, y \in \Omega$, x is said to dominate y if and only if $f_i(x) \leq f_i(y)$ for all $i \in 1, 2, \dots, m$ and $F(x) \neq F(y)$. $x^* \in \Omega$ is Pareto optimal if no $x \in \Omega$ dominates x^* . The set of all the Pareto optimal solutions is called the Pareto set (PS), and the set of their objective vectors is called the Pareto front (PF).

Multiobjective evolutionary algorithms (MOEAs) have been regarded as a major tool for approximating the PF [1]–[7]. In many real-world applications, environmental and measurement noises are inevitable and thus the exact objective function evaluation is impossible. Moreover, noises on different objectives might be of different scales and distributions. In this paper, we assume each function evaluation can only obtain $F(x) + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2 I)$, the deviation σ represents the noise level.

Noises could significantly deteriorate the performance of MOEAs if no extra measures are taken for handling them. Noises can lead to wrong ranking and thus mislead the search. Noises can also make diversity maintaining difficult. Existing approaches for dealing with noises in MOEAs include the following.

- 1) *Ranking*: To reduce the effect of noises, nondominated sorting genetic algorithm-II (NSGA-II)-A [8] adopts α -dominance and uses the support vector machine for establishing a confidence model for ranking. Hypothesis testing [9] and fuzzy theory [10] have also been used to build new dominance relations for noisy MOPs. Probabilistic Pareto ranking [11] uses an error function in estimating the dominance probability between two solutions for noisy MOPs in [9], [12], and [13].
- 2) *Averaging*: It is for de-noising. The most straightforward method is to do objective function evaluation several times independently and then average the obtained value. The deviation of the mean value (i.e., average value) decreases as the number of evaluations increases. Noise-tolerant strength Pareto evolutionary algorithm (SPEA) (NTSPEA) [14] assigns different evaluation numbers to different individual solutions according to their dominance relations in a population. Some researchers (see [15]) also consider the standard deviation of noisy objective function values. To reduce the number of function evaluations, an average of noisy function values of several close solutions can be used as in [16].
- 3) *Modeling*: Noises may affect a single solution significantly. However, their effect on a model can be minimal.

As a solution set can be described as a model, estimation of distribution algorithms (EDAs) have been recently used for dealing with noisy MOPs. For example, restricted Boltzmann machine [17], [18] and univariate marginal distribution algorithm [19] have been adopted for handling noisy MOPs.

The above approaches can deal with noisy MOPs to certain extent, but their performance is not always satisfactory. The existing ranking methods cannot output ranks with a high accuracy for noisy problems. That is the reason why multi-objective probabilistic selection evolutionary algorithm (MOSPEA) [11] with the probabilistic Pareto ranking is not satisfactory on noisy MOPs. The averaging methods need multiple function evaluations for a single solution or similar solutions, which increases their computational cost. Even though some function evaluation saving methods are brought into the averaging-based algorithms (NTSPEA for instance), their efficiency is not high enough. The modeling-based algorithms are more effective than the averaging-based algorithms, but few modeling-based algorithms have considered the characteristics of MOPs.

This paper focuses on modeling approaches for continuous noisy MOPs. Under mild conditions, the PS of a continuous MOP is a piecewise $(m-1)$ -D manifold. Although this regularity property has been successfully used for solving noise-free MOPs [3], it has not yet been considered in noisy MOPs. This paper advocates using this regularity model in MOEAs for dealing with noisy MOPs. Through some analysis in this paper, we argue that the regularity model is efficient for de-noising. We also give the guide on how to use the regularity model in existing MOEAs.

The rest of this paper is organized as follows. Section II introduces the regularity model briefly. Section III analyzes the reason why the regularity model can de-noise efficiently and gives the motivation of this paper. The implementation of the regularity model is shown in Section IV. Section V shows how to use the regularity model in MOEAs. Section VI shows the experimental results and discussions. Finally, the conclusion is given in Section VII.

II. REGULARITY MODEL

Under some mild smoothness conditions, it can be proven by using the well-known Karush–Kuhn–Tucker condition that the PS of a continuous MOP is an $(m-1)$ -D piecewise continuous manifold [20], [21]. This regularity property was first used in MOEAs in [3].

Fig. 1 illustrates a simple way for using the regularity property to obtain a model (called the regularity model) to model a population in MOEAs. Circular points are a population found during the previous search, and the solid curve are the true PS. One can assume that the population scatters around the PS, in other words, the PS can be thought of as a central $(m-1)$ -D manifold of these solutions. To model the PS, one can divide these solutions into several (three in Fig. 1) clusters. Then each cluster is approximated by an $(m-1)$ -D linear model ζ .

Each linear model ζ in one cluster can be defined as (2), it can be obtained by principal component analysis (PCA)

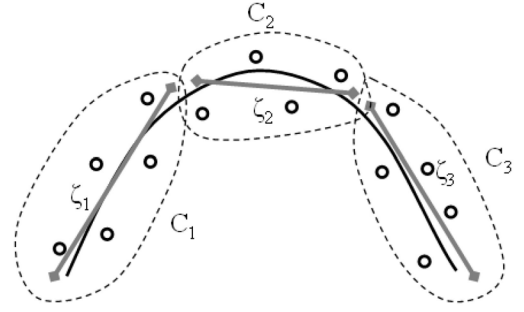


Fig. 1. Modeling from a population by regularity property.

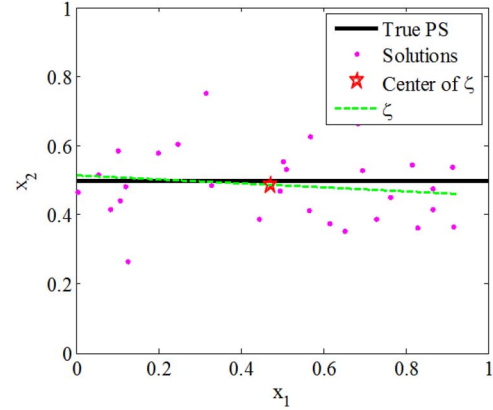


Fig. 2. Illustration of calculation of model ζ .

(the distances from solutions to their projection of ζ are minimized) [3], [22]. In (2), \bar{x} is the center of cluster C , $U^i (1 \leq i \leq n)$ (sorted by the eigenvalue λ^i of the covariance matrix in a descending order) are the principal components of this cluster, θ_i is the free variable in the i th principal subspace

$$\zeta = \bar{x} + \sum_{i=1}^{m-1} \theta_i U^i. \quad (2)$$

An MOP with two decision variables and two objectives in Fig. 2 is used as an example to explain how to obtain ζ . To simplify the problem, its true PS is $x_2 = 0.5$ and 30 solutions around the true PS are classified into one cluster. Firstly, the center \bar{x} of those solutions can be calculated. Then, the principal component U^i can be obtained by PCA. As ζ is $(m-1)$ -D, only the first $m-1$ principal components $U^i (i = 1, \dots, m-1)$ are considered as shown in (2). From the example, \bar{x} and $U^i (i = 1, \dots, m-1)$ construct ζ , the former determines the location of ζ , and the latter determines the shape of ζ .

By the above “modeling” step, an analytical model for PS is obtained. Such a probabilistic model cannot be used directly in population-based MOEAs. Sampling is necessary to sample a population from a model as an operator to generate new solutions in MOEAs.

Fig. 3 illustrates two kinds of sampling from a regularity model. In Fig. 3(a), a population is sampled exactly on the regularity model. In Fig. 3(b), a population is sampled around the regularity model by adding noise ε to ζ , which has been employed in RM-multi-objective estimation of distribution algorithm (MEDA) [3] to add diversity.

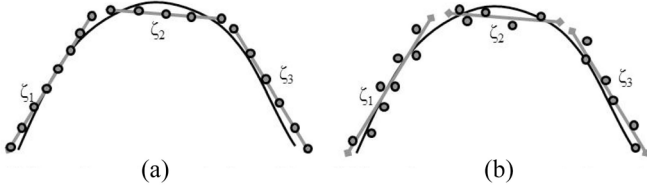


Fig. 3. Sampling from the regularity model. (a) Sampling on the regularity model. (b) Sampling around the regularity model.

The process of sampling from ζ in cluster C can be done by uniformly sampling from free variables θ_i ($1 \leq i \leq m-1$) in (2). The sampling on θ_i should be in the interval $a_i \leq \theta_i \leq b_i$ as (3), where α_i is a uniformly random value in $[0, 1]$, a_i and b_i [shown in (4) and (5)] are the boundaries of sampling in the i th principal subspace of the cluster [3]

$$\Phi = \left\{ x \in R^n | \bar{x} + \sum_{i=1}^{m-1} \theta_i U^i, \theta_i = \alpha_i(b_i - a_i) + a_i \right\} \quad (3)$$

$$a_i = \min_{x \in C} (x - \bar{x})^T U^i \quad (4)$$

$$b_i = \max_{x \in C} (x - \bar{x})^T U^i. \quad (5)$$

III. ANALYSIS AND MOTIVATION

In this section, we consider a major issue which motives us to use the regularity model for dealing with noisy MOPs, that is how well the regularity model can reduce the effect from noises. We first analyze the effect of noises on selection and explain why models can help to reduce noises. Then, through the comparison with other new solution reproduction operators, we study characteristics of the regularity model in noisy environments.

In this paper, we consider the following four different types of MOPs.

- 1) Type I: The objective functions are uni-modal and the mapping F (i.e., the objective function) is nearly symmetrical around the PS. In other words, two symmetry solutions about the PS have the same distance to the PF.
- 2) Type II: The objective functions are multimodal and the mapping F is nearly symmetrical around the PS.
- 3) Type III: The objective functions are uni-modal and the mapping F is not very symmetrical around the PS. In other words, two symmetry solutions about the PS do not have the same distance to the PF.
- 4) Type IV: The objective functions are multimodal and the mapping F is not very symmetrical around the PS.

Table I lists four test instances of different types used in our studies. The PS of F1 and F2 is $0 \leq x_1 \leq 1$, and $x_i = 0.5, i = 2, \dots, n$. The PS of F3 and F4 is $0 \leq x_1 \leq 1$, and $x_i = 0, i = 2, \dots, n$. All these instances are for minimization.

A. Effects of Noises on Selection

To show the effect of noise level (i.e., $\sigma = 0, \sigma = 0.1$, and $\sigma = 0.2$) on selection, we conduct the following experiments on F1–F4 with two decision variables.

- 1) Evenly generate 51×51 solutions in $[0, 1]^2$ as an initial population Pop_i from the search space.

TABLE I
TEST INSTANCES

No.	Variables	Objectives	Type
F1	$x \in [0, 1]^n$	$f_1 = (x_1)g$ $f_2 = (1 - x_1)g$ $g = \sum_{i=2}^n (x_i - 0.5)^2$	I
F2	$x \in [0, 1]^n$	$f_1 = (x_1)g$ $f_2 = (1 - x_1)g$ $g = \sum_{i=2}^n (x_i - 0.5)^2 - \cos(10\pi(x_i - 0.5))$	II
F3	$x \in [0, 1]^n$	$f_1 = (x_1)g$ $f_2 = (1 - x_1)g$ $g = \sum_{i=2}^n x_i^2$	III
F4	$x \in [0, 1]^n$	$f_1 = (x_1)g$ $f_2 = (1 - x_1)g$ $g = \sum_{i=2}^n x_i^2 - \cos(10\pi x_i)$	IV

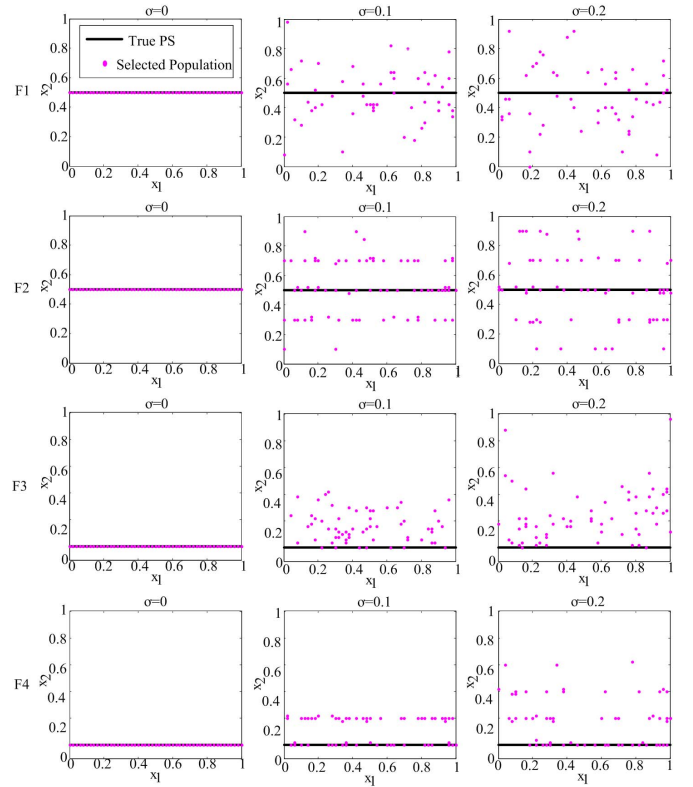


Fig. 4. Selected population Pop_s on F1–F4 with different noise levels ($\sigma = 0, \sigma = 0.1$, and $\sigma = 0.2$).

- 2) Use a nondominated sort (e.g., fast nondominated rank sort [1], nondominated rank sort [23], deductive sort [24], corner sort [25], and efficient nondominated sort [26]) to select 51 solutions as the selected population Pop_s .

Fig. 4 presents the experimental results. All the solutions in the selected population Pop_s are very close to the true PS in the noise-free case (i.e., $\sigma = 0$). The deviation of the points in the selected population is increasing as σ grows. Solutions closer to the true PS have more chances to generate optimal solutions in the future generations. They can promote the evolutionary process and be viewed as good solutions. However, good solutions cannot be selected due to noises in the

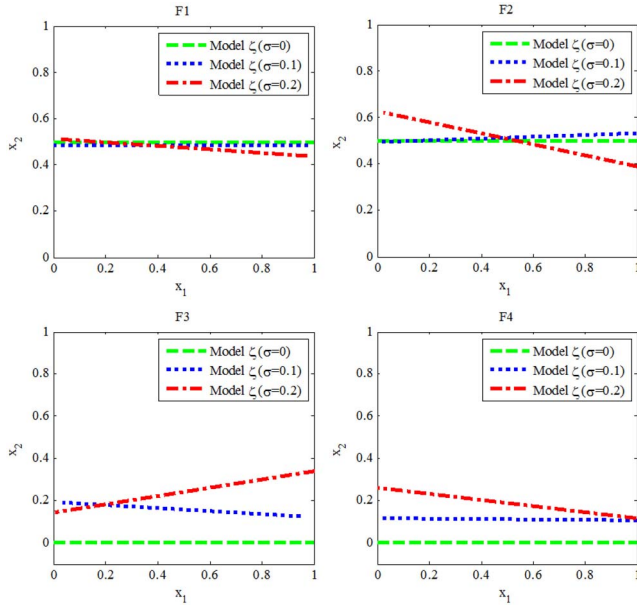


Fig. 5. Model ζ restored from the selected population Pop_s on F1–F4 with different noise levels ($\sigma = 0$, $\sigma = 0.1$, and $\sigma = 0.2$).

objective space. It is clear that the effect of noises varies from instance to instance.

A majority of selected solutions on uni-modal F1 and F3 (types I and III) are close to the PS. In contrast, quite a number of selected solutions on multimodal F2 and F4 (types II and IV) are close to some local optima.

On F1 and F2 (types I and II), selected solutions are distributed on both sides of the PS. In contrast, on F3 and F4 (types III and IV), most solutions are distributed on one side of the PS due to their less-symmetrical mappings around the true PS.

B. Models for Selected Population

Now, we study if one can build a good model for the PS from the selected solutions. We do PCA on these solutions to build a linear model for modeling their distributions. The results are plotted in Fig. 5.

Fig. 5 presents the experimental results. In the noise-free case, the obtained model approximates the PS very much. This is not surprising because the selected solutions are very close to the PS. As the noise level σ increases, the model quality is getting worse.

On F1 and F2 (types I and II) with symmetrical mapping, the model can approximate to the true PS reasonably good, particularly when σ is small. However, models are very poor on F3 and F4 even when σ is small.

These experimental results suggest that models are useful for noisy MOPs of types I and II, but are not effective for types III and IV. The symmetrical degree of the mapping F affects the de-noising performance of the regularity model. There are two kinds of F in those test instances, they are both extreme cases. F1 and F2 have the perfectly symmetrical mapping and F3 and F4 have the completely nonsymmetrical mapping. In practice, the symmetrical degree of most

problems is between these two extremes. Therefore, the results of F1 and F2 show the best performance of the regularity model for noisy MOPs, whereas the results of F3 and F4 show the worst performance of regularity model for noisy MOPs.

C. Efficiency of Regularity Model

The efficiency of a de-noising approach means how well it can guide the population to approximate the true PS, which can be quantitatively measured by the average distance to the true PS in the decision space. That distance can be calculated by averaging the shortest distances of Pop_s to a set of uniform samples from true PS.

Now, we study how well a regularity model works in noisy environments by comparing with EDA (a different model-based approach that has been used for de-noising). We use the EDA with the univariate marginal product model [19] as a compared approach. SBX [27] without any de-noising ability is also compared as a reference. Therefore, we conduct the following experiments on F1–F4 with 2–30 decision variables in two different noise levels ($\sigma = 0.1$ and $\sigma = 0.2$).

- 1) Randomly generate 500 solutions from the search space.
- 2) Find the nondominated solutions as the selected population Pop_s .
- 3) Build a regularity model from Pop_s as the modeling step in [3], and sample 100 solutions exactly on the obtained model as in Fig. 3(a).
- 4) Build a univariate marginal product model [19] from Pop_s , and sample 100 solutions on the obtained model.
- 5) Use SBX [27] to generate 100 solutions from Pop_s .

These compared approaches have different output. The regularity model outputs analytical models (ζ with \bar{x} and $U^i (i = 1, \dots, m-1)$). Whereas, EDA and SBX produce solutions. Therefore, we use solutions sampled from the obtained regularity model as in Fig. 3(a) for a fair comparison with EDA and SBX.

Fig. 6 presents the average distance to the true PS of the 100 solutions generated by each approach on the four test instances with different numbers of decision variables under two different noise levels ($\sigma = 0.1$ and $\sigma = 0.2$). The average distance of the solutions in Pop_s to the true PS is also presented in this figure as a reference. It is clear that on the instances with two or three decision variables, all these approaches perform similarly. Their obtained solutions have about the same average distance to the true PS. However, on all the four instances with more than four decision variables, the average distance of new solutions generated by the regular model to the true PS is smaller than that of the solutions in Pop_s . Recalling that Pop_s was selected from noisy environments, this suggests that the regularity model does have a de-noising ability. In contrast, the average distance of the new solutions generated by SBX is about same as that of the solutions in Pop_s . Therefore, SBX is unable to de-noise. The EDA approach can de-noise on F3 and F4 with a large number of decision variables, but it is not as good as the regular model approach. A major reason is that the EDA approach does not make use of the PS regularity property. With the PS regularity property, the dimension of the model is reduced from n to $m-1$. Therefore, when n

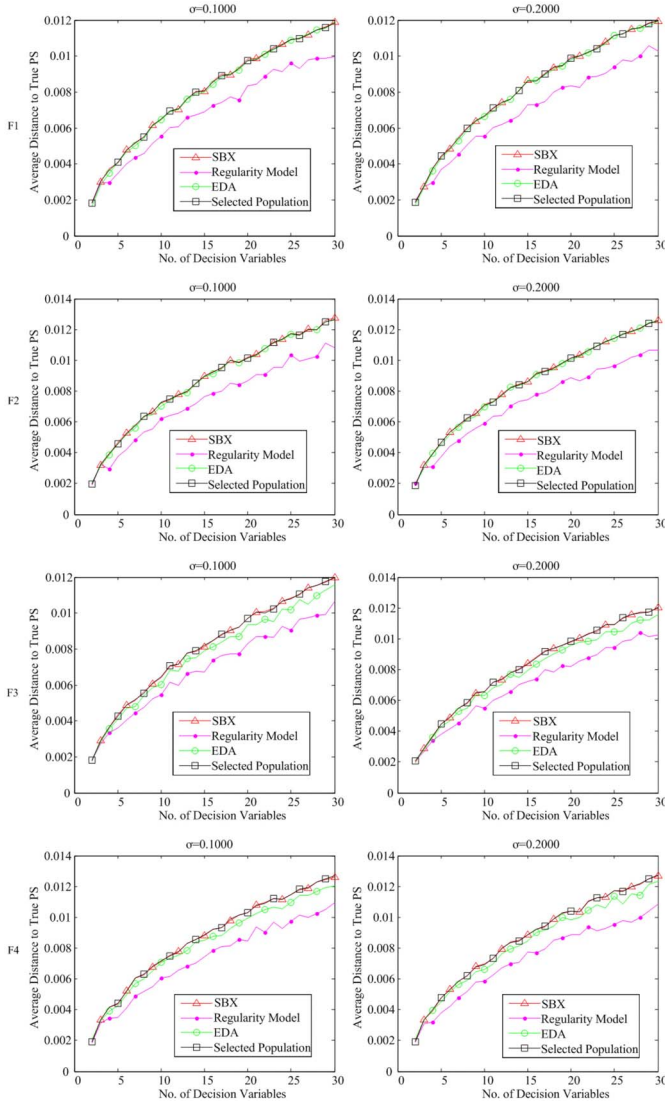


Fig. 6. Average distance of the selected population Pop_s and samplings obtained by SBX, EDA, and the regularity model to the true PS on F1–F4 with 2–30 decision variables in different sizes of noises ($\sigma = 0.1$ and $\sigma = 0.2$).

is larger than $m - 1$, which often happens in MOPs, the regularity model, as a de-noising tool, is more effective than the EDA approach.

D. Motivation

With advantages of models and regularity property, the regularity model is more effective than other model-based methods in de-noising. Having this in mind, we believe that a regularity model on the obtained nondominated solution set can help existing MOEAs for noise-free problems to deal with noisy MOPs. Therefore, we will focus on how the regularity model works for noisy problems in Section IV and how to use the regularity model for improving existing MOEAs in Section V.

IV. PROPOSED REGULARITY MODEL FOR DE-NOISING

A simple implementation of the regularity model has been proposed for noise-free MOPs in [3]. In this section, we modify it for noisy MOPs.

Algorithm 1 Pseudo Code of the Self-Adaptive Local PCA

- 1: **Parameters:** Pop —the non-domination solution set in the current population, N —the size of Pop , m —No. of objectives.
- 2: **For** $K = 1 : \lfloor N/m \rfloor$
- 3: Divide Pop into K clusters by local PCA [3], [22].
- 4: **If** $\frac{\sum_{i=1}^{m-1} \lambda_j^i}{\sum_{i=1}^n \lambda_j^i} > P_0, 1 \leq j \leq K$
- 5: Break.
- 6: **End**
- 7: **End**

A. Modeling

Local PCA [22] rather than K -means [28] is employed to classify the population into K clusters (i.e., K linear models), because it suits the task of manifold division. Considering a PS with K segments of manifold, each cluster C_j contains one $(m - 1)$ -D manifold, i.e., the contribution of principal components should concentrate on the first $(m - 1)$ components as (6), where λ_j^i is the i th eigenvalue in C_j and P_0 is one parameter of PCA. P_0 is usually set as 0.7, 0.8, and 0.9 [29], we set it as 0.9 in this paper

$$\frac{\sum_{i=1}^{m-1} \lambda_j^i}{\sum_{i=1}^n \lambda_j^i} \geq P_0. \quad (6)$$

However, the cluster method with a fixed number of clusters K cannot suit all different distributions well. In view of this, we design a self-adaptive method in Algorithm 1. The self-adaptive local PCA tries to cluster the population with an increasing K and stops until all the clusters satisfy (6) [the population is divided into $K(m - 1)$ -D manifolds].

The minimal cluster size to build an $(m - 1)$ -D linear model is m . That is the reason why we iterate K from 1 to $\lfloor N/m \rfloor$ in Algorithm 1. PS is a piecewise $(m - 1)$ -D manifold, there would not be as many as $\lfloor N/m \rfloor$ segments in most cases. Our method stops once $K(m - 1)$ -D manifolds are found. Therefore, it does not incur any heavy computational cost.

B. Sampling

Aiming at noisy MOPs, we improve the sampling step especially in two aspects [endpoint maintenance and uniform sampling matrix (USM)].

1) *Endpoint Maintenance:* In the regularity model, only the first $(m - 1)$ principal components $U^i (i = 1, \dots, m - 1)$ of ζ are considered. We define that either point x_{\min}^i or x_{\max}^i [shown in (7) and (8)] is an endpoint P in the decision space, which contributes to the spread of PS. Because of noises, the solution set is very changeable in different generations. That is the reason why endpoints (the spread of PS) are hard to be maintained

$$x_{\min}^i = \arg \min_{x \in C} (x - \bar{x})^T U^i, i = 1, \dots, m - 1 \quad (7)$$

$$x_{\max}^i = \arg \max_{x \in C} (x - \bar{x})^T U^i, i = 1, \dots, m - 1. \quad (8)$$

There are two cases between two generations, either model ζ changes significantly or slightly. When ζ changes significantly, the location and shape of ζ change significantly,

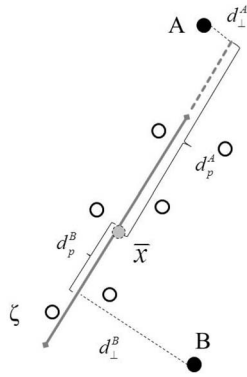


Fig. 7. Illustration of the endpoint maintenance strategy.

the endpoints in the previous generation may not be on the new ζ , thus they do not need to be maintained. However, when ζ changes slightly, the location and shape of ζ change slightly but the spread might be changeable, the endpoints in the previous generation may be on the new ζ , they need to be kept in the new ζ to maximize the spread of PS.

The changing degree of ζ between generations can be quantitatively calculated. Taking Fig. 7 as an example, endpoints A and B are two endpoints in the previous generation, ζ is the model in this generation. Determining whether those two points are on ζ can be done by their projection to ζ . We find the projection distance d_p^A of A is much longer than its offset distance d_\perp^A , thus A is on ζ . By adding A, the spread of ζ increases. However, the situation of B is different, thus B is not on ζ . Therefore, the ratio $d_p^2/(d_p^2 + d_\perp^2)$ of one endpoint x_e shows the changing degree, when the ratio is higher than 0.95, ζ changes slightly, x_e needs to be maintained. Distances d_p and d_\perp of x_e can be calculated as shown below

$$d_p = \sqrt{\sum_{i=1}^{m-1} [(x_e - \bar{x})^T U^i]^2} \quad (9)$$

$$d_\perp = \sqrt{\sum_{i=m}^n [(x_e - \bar{x})^T U^i]^2}. \quad (10)$$

We record endpoints in every generation to maintain endpoints as Algorithm 2. When an endpoint from the previous generation is on ζ , the model changes slightly between generations, the points need to be maintained for the calculation of the points x_{\min}^i and x_{\max}^i for the t th generation by (7) and (8). Thus, the spread of PS can be kept well between different generations.

2) *Uniform Sampling Matrix*: As shown in (3), the sampling step is to sample uniformly in an $(m-1)$ -D cubic space. Uniformly sampling from the regularity model can provide better diversity for the population. We randomly sample a matrix $\text{rand}(N, m-1)$ in every generation and accumulate these matrices to form a relatively USM. Our algorithm adds new random samples and deletes extra samples in every generation to form a more USM. Some diversity maintenance strategies in MOEAs can achieve the deleting task, such as the crowding distance in NSGA-II [1], environment selection in SPEA2 [4], and the harmonic distance in [30].

Algorithm 2 Pseudo Code of the Endpoint Maintenance for an Obtained Model ζ in the t th Generation

- 1: **Parameters**: P_{t-1}^l -the l -th endpoint in the $t-1$ -th generation, Pop -the solution set for building the model, and ζ -the obtained model in the t -th generation
- 2: **For** each P_{t-1}^l
- 3: **If** P_{t-1}^l is on ζ
- 4: Add P_{t-1}^l to Pop .
- 5: **End**
- 6: **End**
- 7: Calculate a_i and b_i as Equations (4) and (5), where $x \in Pop$, $i = 1, \dots, m-1$.
- 8: Add the points x_{\min}^i and x_{\max}^i (shown in Equations (7) and (8)) to P_t for the t -th generation.

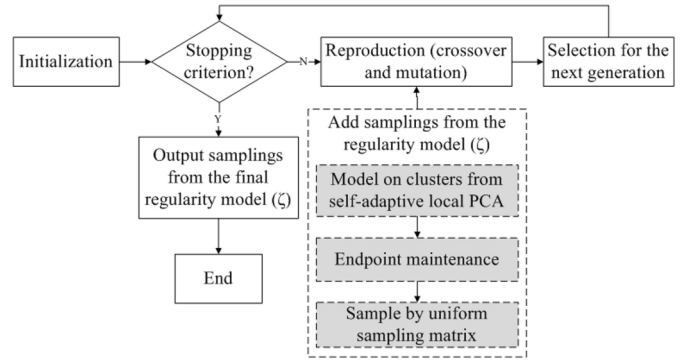


Fig. 8. Flow-chart of embedding the regularity model in MOEAs, the solid line represents the general flow of MOEAs, and the dotted line represents the flow of building regularity model on the nondominated solution set (modeling after the clustering by self-adaptive local PCA, endpoint maintenance, and sampling by USM).

V. IMPROVE EXISTING MOEAs BY REGULARITY MODEL

MOEAs have been developed to handle different MOPs. However, existing MOEAs for noisy problems have not taken the full use of MOEAs for different types of MOPs. Therefore, the main purpose of this paper is to improve the performance of any existing MOEAs on a wide range of MOPs with noises by adding the regularity model as a part of them. To show how to embed the regularity model in existing MOEAs, we use NSGA-II [1] as an example in the following sections.

A. Example MOEA Embedded With Regularity Model

In Section II, we have shown that the regularity model can efficiently reduce noises in the population. Building a regularity model for the nondominated solution set in every generation can help MOEAs to de-noise. Thus, the nondominated solution set is modeled by a regularity model in every generation. To be compatible with population-based MOEAs, the regularity model samples a population from a model as shown in Section IV-B. The final output is a set of samplings from the obtained regularity model in the last generation.

As shown in Fig. 8, the regularity model can be embedded in existing MOEAs as an additional reproduction operator, which is similar to that of [31], the other parts of the MOEA are not affected by the regularity model. In this paper, we embed the

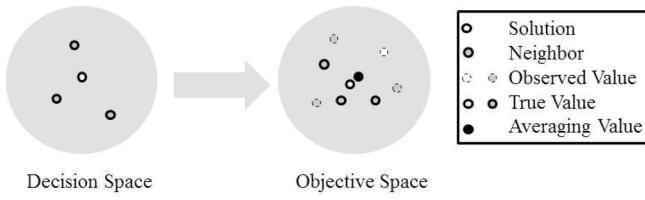


Fig. 9. Illustration of the ED strategy.

regularity model in NSGA-II (called RM-NSGA-II) for the following experiment to study the improvement of MOEAs by the regularity model on noisy MOPs, because NSGA-II is the most widely-studied MOEA. Its characteristics are well-known, thus our idea can be understood easily.

B. Extra De-Noising

To provide a more precise sample points for the regularity model, we add a preprocessing step [an extra de-noising (ED) strategy] to RM-NSGA-II. Inspired by extended averaging [16], similar individuals in the decision space can be viewed as of approximated objective values. As shown in Fig. 9, the observed objective values are far from their true objective values due to noises, but the average of neighbors (gray dots) in the decision space is approximated to the true objective value.

Therefore, there is a set to record the solutions during the optimization process. After observing the objectives of a new solution, its decision variables are compared with the existing solutions in the set to check if any similar solutions have been searched before. If there are any similar solutions in the set, the averages of both decision variables and objective values are employed for the solution after the extra strategy. We use the L_1 -norm-based distance as the similarity evaluation.

VI. COMPUTATIONAL EXPERIMENTS

A. Test Problems, Evaluation Metrics, and Parameter Settings

In order to evaluate our new idea for noisy MOPs, we employ the ZDT [32], DTLZ [33], and WFG [34] problems with different noise levels as the test problems in our experiments. The details of those test problems are shown in Table II. We adopt generational distance (GD) [35], minimal spacing [36], and inverted generational distance (IGD) [37] to evaluate the performance of different algorithms. GD is the average distance from the obtained PF to the true PF, which describes the convergence of the obtained PF. Minimal spacing is a metric for uniformity, which uses nonduplicated distances for its final calculation. IGD is the average distance from the true PF to the obtained PF, whose value reflects both convergence and diversity of the obtained PF. In the following experiments, the stopping criterion is set as 50 000 function evaluations, and the reproduction methods in the compared algorithms are set as SBX ($\eta = 15$ and probability = 1) and polynomial mutation ($\eta = 15$ and probability = 0.1). All the experiments are repeated independently for 30 times.

TABLE II
CHARACTERISTICS OF TEST PROBLEMS

Problem	m	n	Note
ZDT1	2	30	uni-modal, convex PF, continues PS
ZDT2	2	30	uni-modal, concave PF, continues PS
ZDT3	2	30	uni-modal, disconnected PF, disconnected PS
ZDT4	2	10	multi-modal, convex PF, continues PS
DTLZ1	3	7	multi-modal, convex PF, continues PS
DTLZ2	3	12	uni-modal, convex PF, continues PS
DTLZ3	3	7	multi-modal, convex PF, continues PS
DTLZ4	3	12	uni-modal, convex PF, continues PS
WFG1	2	20	uni-modal, nonseparable, continues PS
WFG2	2	20	multi-modal, disconnected PF, continues PS
WFG3	2	20	multi-modal, continues PS
WFG4	2	20	multi-modal, concave PF, continues PS
WFG5	2	20	deceptive, concave PF, continues PS
WFG6	2	20	concave PF, continues PS
WFG7	2	20	uni-modal, nonseparable, concave PF, continues PS
WFG8	2	20	concave PF, continues PS
WFG9	2	20	multi-modal, deceptive, concave PF, continues PS

B. Experiments for Extra De-Noising

The ED strategy is a special noise handling operation in RM-NSGA-II, whose effect to RM-NSGA-II is analyzed in this section. We compare the proposed algorithms with and without the ED strategy on ZDT2 and DTLZ2. The results on GD, minimal spacing, and IGD are shown in Table III, which are analyzed by Wilcoxon signed-rank test [38]. For the noise-free problems, the algorithm without the ED strategy has better performance than that with the ED strategy. This is because that the averaging in neighborhoods for the noise-free problems adds uncertainty to function observations. However, for the noisy problems, the ED strategy helps the algorithm to build an accurate model.

From the results in Table III, we find that the ED strategy improves the performance of RM-NSGA-II for noisy MOPs on both convergence and diversity. Although the improvement on the two-objective problem ZDT2 is small, the improvement on the three-objective problem DTLZ2 is significant. In short, the ED strategy can help the proposed algorithm to obtain a better model in noisy environments.

C. Experiments for Endpoint Maintenance

In this section, we analyze the effect of the endpoint maintenance strategy in the regularity model. We compare the proposed algorithms with and without the endpoint maintenance strategy on ZDT2 and DTLZ2 ($\sigma = 0.1$ and $\sigma = 0.2$). The distances to the true endpoints of the proposed algorithms with and without the endpoint maintenance strategy over generations are shown in Fig. 10. With the endpoint maintenance strategy, the proposed algorithm can obtain the solution close to the true endpoints in a noisy environment, which leads to a larger PS spread than that without the endpoint maintenance strategy.

D. Experiments for Uniform Sampling Matrix

In this section, we analyze the USM in the regularity model. We compare the proposed algorithms with and without the USM on ZDT2 and DTLZ2. The results on GD, minimal

TABLE III
RESULTS OF THE PROPOSED ALGORITHMS WITH AND WITHOUT THE ED STRATEGY ANALYZED BY WILCOXON SIGNED-RANK TEST. THE SIGNIFICANT RESULTS ARE IN BOLD FACE (SIGNIFICANCE LEVEL = 0.05)

	σ	GD		Minimal Spacing		IGD	
		Algorithm with ED	Algorithm without ED	Algorithm with ED	Algorithm without ED	Algorithm with ED	Algorithm without ED
ZDT2	0.00	0.0011 \pm 0.0002	0.0001\pm0.0000	0.0062 \pm 0.0010	0.0041\pm0.0022	0.0132\pm0.0143	0.1401 \pm 0.1964
	0.05	0.0029 \pm 0.0004	0.0146 \pm 0.0406	0.0069 \pm 0.0006	0.0119 \pm 0.0125	0.0154 \pm 0.0018	0.0071\pm0.0024
	0.10	0.0050 \pm 0.0009	0.0121 \pm 0.0222	0.0064\pm0.0013	0.0172 \pm 0.0266	0.0452 \pm 0.1096	0.0106\pm0.0060
	0.20	0.0078 \pm 0.0018	0.0058\pm0.0116	0.0058\pm0.0018	0.0070 \pm 0.0025	0.1166 \pm 0.2045	0.0257\pm0.0204
DTLZ2	0.00	0.0288 \pm 0.0242	0.0116\pm0.0266	0.1275 \pm 0.0677	0.0816\pm0.0499	0.1462 \pm 0.0852	0.1109 \pm 0.0648
	0.05	0.0068\pm0.0135	0.0074 \pm 0.0067	0.0963\pm0.0375	0.1807 \pm 0.0623	0.0607\pm0.0091	0.1205 \pm 0.0211
	0.10	0.0018\pm0.0002	0.0074 \pm 0.0027	0.0855\pm0.0102	0.1847 \pm 0.0678	0.0583\pm0.0044	0.1530 \pm 0.0273
	0.20	0.0026\pm0.0011	0.0143 \pm 0.0048	0.0902\pm0.0069	0.2014 \pm 0.0573	0.0675\pm0.0083	0.2392 \pm 0.0593

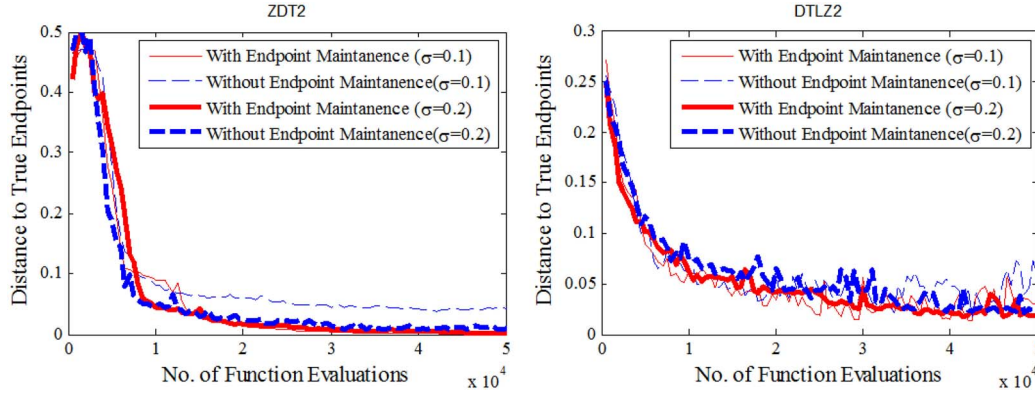


Fig. 10. Distances to the true endpoints of the proposed algorithms with and without the endpoint maintenance strategy over generations on ZDT2 and DTLZ2 ($\sigma = 0.1$ and $\sigma = 0.2$).

TABLE IV
RESULTS OF THE PROPOSED ALGORITHMS WITH AND WITHOUT THE USM ANALYZED BY WILCOXON SIGNED-RANK TEST. THE SIGNIFICANT RESULTS ARE IN BOLD FACE (SIGNIFICANCE LEVEL = 0.05)

	σ	GD		Minimal Spacing		IGD	
		Algorithm with USM	Algorithm without USM	Algorithm with USM	Algorithm without USM	Algorithm with USM	Algorithm without USM
ZDT2	0.00	0.0011 \pm 0.0002	0.0011 \pm 0.0002	0.0061\pm0.0010	0.0144 \pm 0.0021	0.0132\pm0.0143	0.0180 \pm 0.0182
	0.05	0.0029 \pm 0.0004	0.0028 \pm 0.0003	0.0069\pm0.0006	0.0153 \pm 0.0017	0.0154\pm0.0018	0.0173 \pm 0.0015
	0.10	0.0050 \pm 0.0009	0.0049 \pm 0.0007	0.0064\pm0.0013	0.0148 \pm 0.0018	0.0452 \pm 0.1096	0.0268 \pm 0.0032
	0.20	0.0078 \pm 0.0018	0.0074 \pm 0.0012	0.0058\pm0.0018	0.0145 \pm 0.0029	0.1166 \pm 0.2045	0.0585 \pm 0.1085
DTLZ2	0.00	0.0288 \pm 0.0242	0.0176 \pm 0.0295	0.1275 \pm 0.0677	0.1002\pm0.0582	0.1462 \pm 0.0852	0.1730 \pm 0.1074
	0.05	0.0068 \pm 0.0135	0.0058 \pm 0.0101	0.0963 \pm 0.0375	0.1037 \pm 0.0285	0.0607\pm0.0091	0.0701 \pm 0.0080
	0.10	0.0018 \pm 0.0002	0.0040 \pm 0.0084	0.0855\pm0.0102	0.1036 \pm 0.0268	0.0583\pm0.0044	0.0709 \pm 0.0079
	0.20	0.0026 \pm 0.0011	0.0026 \pm 0.0008	0.0902 \pm 0.0069	0.0916 \pm 0.0097	0.0675\pm0.0083	0.0807 \pm 0.0247

spacing, and IGD are shown in Table IV, which are analyzed by Wilcoxon signed-rank test [38]. As the USM has no relation to the modeling in RM-NSGA-II, GD is not influenced by the USM. For minimal spacing, the improvement is significant especially on ZDT2. The reason is that the USM becomes more and more uniform by generations than a random sampling matrix. Additionally, the value of IGD is significantly improved on both ZDT2 and DTLZ2 because of better diversity from the USM.

E. Comparative Experiments

In order to test the performance of RM-NSGA-II on noisy MOPs, we employ several different MOEAs as compared algorithms in our experiments (shown in Table V). They are NSGA-II, which is one of the most well-known MOEAs and the base of RM-NSGA-II, RM-MEDA, which is a regularity

TABLE V
EXPLANATIONS FOR COMPARED ALGORITHMS

Algorithm	Aim	Note
NSGA-II	Non-noisy MOPs	A well-known MOEA with fast non-dominated sort and crowding distance
RM-MEDA	Non-noisy MOPs	An MOEA with regularity model
MOSPEA	Noisy MOPs	A SPEA with probabilistic Pareto ranking
NTSPEA	Noisy MOPs	A SPEA with different arrangement (No. of function evaluations) according to dominance relations

model-based MOEA, MOSPEA, which is a probabilistic Pareto ranking-based MOEA for noisy MOPs, and NTSPEA, which is an averaging-based MOEA for noisy MOPs. We conduct the experiments on the ZDT, DTLZ, and WFG problems. All the experiments are terminated after 50 000 function evaluations and repeated for 30 independent runs. Other parameter

TABLE VI
IGD VALUES OF RM-NSGA-II, NSGA-II, RM-MEDA, MOSPEA, AND NTSPEA ANALYZED BY WILCOXON SIGNED-RANK TEST ON THE ZDT PROBLEMS. THE SIGNIFICANT RESULTS ARE IN BOLD FACE (SIGNIFICANCE LEVEL = 0.05)

	σ	RM-NSGA-II	NSGA-II	RM-MEDA	MOSPEA	NTSPEA
ZDT1	0.00	0.0095±0.0072	0.0039±0.0000	0.0038±0.0000	0.1923±0.1626	0.1201±0.0508
	0.05	0.0122±0.0012	0.0300±0.0056	0.0319±0.0046	0.0267±0.0064	0.0325±0.0101
	0.10	0.0184±0.0030	0.0617±0.0538	0.0530±0.0112	0.0421±0.0103	0.0524±0.0099
	0.20	0.0302±0.0084	0.0847±0.0211	0.0840±0.0148	0.0579±0.0191	0.0899±0.0212
ZDT2	0.00	0.0132±0.0143	0.0039±0.0000	0.0038±0.0000	0.2265±0.1704	0.2743±0.1535
	0.05	0.0154±0.0018	0.0480±0.0079	0.0506±0.0107	0.0534±0.0129	0.0514±0.0108
	0.10	0.0452±0.1096	0.0940±0.0208	0.0937±0.0166	0.0948±0.0223	0.1007±0.0213
	0.20	0.1167±0.2046	0.2051±0.0482	0.1822±0.0502	0.1895±0.0335	0.1971±0.0423
ZDT3	0.00	0.0177±0.0051	0.0050±0.0014	0.0047±0.0002	0.4862±0.1434	0.1465±0.0524
	0.05	0.0202±0.0018	0.0616±0.0106	0.0483±0.0068	0.0616±0.0121	0.0610±0.0096
	0.10	0.0230±0.0059	0.0773±0.0227	0.0559±0.0110	0.0778±0.0282	0.0868±0.0243
	0.20	0.0290±0.0040	0.1189±0.0294	0.1168±0.0300	0.0990±0.0305	0.1271±0.0313
ZDT4	0.00	0.0337±0.0355	0.0039±0.0000	0.0039±0.0000	0.3335±0.2664	0.2045±0.1687
	0.05	0.0521±0.0528	0.0265±0.0096	0.0353±0.0306	0.0233±0.0079	0.0327±0.0135
	0.10	0.0836±0.1609	0.0532±0.0169	0.0499±0.0111	0.0353±0.0126	0.0532±0.0101
	0.20	0.0831±0.1219	0.0768±0.0298	0.0843±0.0309	0.0556±0.0276	0.0877±0.0343

TABLE VII
IGD VALUES OF RM-NSGA-II, NSGA-II, RM-MEDA, MOSPEA, AND NTSPEA ANALYZED BY WILCOXON SIGNED-RANK TEST ON THE DTLZ PROBLEMS. THE SIGNIFICANT RESULTS ARE IN BOLD FACE (SIGNIFICANCE LEVEL = 0.05)

	σ	RM-NSGA-II	NSGA-II	RM-MEDA	MOSPEA	NTSPEA
DTLZ1	0.00	16.8969±21.8599	0.0396±0.0014	0.4062±0.3526	0.9794±0.9229	1.2366±0.9665
	0.05	5.6881±10.9509	0.4921±0.4832	1.0271±0.9588	0.0957±0.0963	1.5826±0.8065
	0.10	5.1482±18.8805	0.4726±0.4419	0.8730±0.7300	0.1247±0.0273	1.2291±1.0801
	0.20	3.1770±5.6823	0.5204±0.4148	1.5693±0.9887	0.2872±0.1868	1.8786±1.2472
DTLZ2	0.00	0.1462±0.0852	0.0521±0.0011	0.0640±0.0019	0.2610±0.0694	0.0672±0.0036
	0.05	0.0608±0.0094	0.1001±0.0061	0.1148±0.0071	0.1404±0.0088	0.1351±0.0116
	0.10	0.0584±0.0044	0.1655±0.0169	0.1636±0.0143	0.2363±0.0275	0.1893±0.0183
	0.20	0.0678±0.0085	0.3097±0.0417	0.2809±0.0291	0.3979±0.0219	0.3321±0.0449
DTLZ3	0.00	72.8487±71.6040	0.0530±0.0034	0.5137±0.5572	0.7629±0.9064	1.2291±0.9858
	0.05	36.6895±51.0523	0.6300±0.5441	1.4291±0.7777	0.2691±0.1142	1.6443±1.6503
	0.10	13.5323±19.9786	0.7002±0.5621	1.5623±1.4506	0.3414±0.0838	1.7944±1.2928
	0.20	26.6722±65.3947	0.8297±0.4108	1.9540±1.4630	0.4984±0.2608	1.3951±0.9785
DTLZ4	0.00	0.4330±0.2999	0.3473±0.3270	0.0767±0.0683	0.3963±0.2684	0.2381±0.2814
	0.05	0.1108±0.0058	0.1502±0.1580	0.3132±0.1656	0.1968±0.1521	0.1512±0.0829
	0.10	0.1598±0.1136	0.2569±0.2192	0.1774±0.0160	0.2448±0.1128	0.2568±0.1465
	0.20	0.3358±0.0441	0.3348±0.1165	0.3524±0.2026	0.4506±0.1281	0.3386±0.0564

settings have been shown in Section VI-A. The comparative results are shown in Tables VI–VIII.

1) *Results:* Table VI shows the IGD values of compared algorithms on the ZDT problems. For the noise-free ZDT problems, NSGA-II and RM-MEDA can outperform other algorithms. For noisy ZDT1-3, RM-NSGA-II outperforms other compared algorithms, but for noisy ZDT4, RM-NSGA-II cannot perform better than the MOEAs for noisy problems (NTSPEA and MOSPEA). As ZDT4 is a multimodal problem, noises make algorithms easily trapped in local optima. Thus, it is hard for the regularity model to learn to jump out of local optima due to its low exploration ability. In contrast, NTSPEA and MOSPEA have strategies to improve the convergence for noisy MOPs. That is the reason why RM-NSGA-II has poor performance on ZDT4, which is similar to the case of F4 in Section III-C.

Table VII shows the IGD values of compared algorithms on the DTLZ problems. For noise-free DTLZ problems, both NSGA-II and RM-MEDA are the winners. The results of noisy DTLZ problems are similar to that of ZDT. RM-NSGA-II can outperform other compared algorithms on noisy DTLZ2 and DTLZ4, but cannot outperform other

compared algorithms on noisy DTLZ1 and DTLZ3 that are multimodal, because the regularity model concentrates on local optima. The reason of the less satisfactory IGD of RM-NSGA-II on DTLZ4 comes from poor diversity. As we know, the mapping relation of DTLZ4 from PS to PF is not uniform, but RM-NSGA-II samples uniformly on PS, which leads to poor diversity on PF.

Table VIII shows the IGD values of compared algorithms on the WFG problems. For the noise-free WFG problems, NSGA-II outperforms other algorithms on WFG1, WFG4, WFG5, and WFG8, RM-MEDA outperforms other algorithms on WFG2 and WFG7. RM-NSGA-II outperforms other algorithms on most noisy WFG problems except for WFG2 and WFG7, where RM-MEDA is the winner.

To explore the limitation of the noise level that compared algorithms can deal with, we test these five algorithms on ZDT2 and DTLZ2 with higher level noises (up to $\sigma = 0.5$) for 30 independent runs. The curves of IGD value versus the σ value are shown in Fig. 11.

All the compared algorithms increase their IGD values as the σ value increases, and RM-NSGA-II has the smallest

TABLE VIII
IGD VALUES OF RM-NSGA-II, NSGA-II, RM-MEDA, MOSPEA, AND NTSPEA ANALYZED BY WILCOXON SIGNED-RANK TEST ON THE WFG PROBLEMS. THE SIGNIFICANT RESULTS ARE IN BOLD FACE (SIGNIFICANCE LEVEL = 0.05)

	σ	RM-NSGA-II	NSGA-II	RM-MEDA	MOSPEA	NTSPEA
WFG1	0.00	0.9580±0.0730	0.7769±0.0182	1.0909±0.0089	1.3596±0.0692	0.9242±0.0300
	0.05	1.2230±0.0240	1.2638±0.0110	1.2621±0.0052	1.2611±0.0114	1.2646±0.0083
	0.10	1.2827±0.0190	1.2809±0.0061	1.2751±0.0038	1.2752±0.0072	1.2800±0.0052
	0.20	1.3239±0.0345	1.2876±0.0057	1.2812±0.0048	1.2811±0.0039	1.2893±0.0068
WFG2	0.00	1.5592±0.0707	1.5140±0.0491	1.4410±0.0018	1.9927±0.2205	1.5567±0.0466
	0.05	1.5833±0.0533	1.5971±0.0478	1.5409±0.0324	1.6215±0.0470	1.6159±0.0542
	0.10	1.5936±0.0537	1.6466±0.0466	1.5789±0.0251	1.6475±0.0688	1.6585±0.0592
	0.20	1.7222±0.1148	1.7337±0.0984	1.7292±0.0769	1.7716±0.0771	1.7726±0.0926
WFG3	0.00	0.2460±0.0808	0.0236±0.0067	0.0241±0.0032	0.5307±0.0904	0.1098±0.0254
	0.05	0.0944±0.0213	0.1056±0.0214	0.0985±0.0187	0.1126±0.0225	0.1294±0.0228
	0.10	0.1330±0.0382	0.1709±0.0298	0.1689±0.0221	0.1464±0.0275	0.1776±0.0234
	0.20	0.1900±0.0582	0.2608±0.0534	0.2990±0.0408	0.2129±0.0387	0.2729±0.0379
WFG4	0.00	0.2569±0.0887	0.0163±0.0022	0.0318±0.0045	0.6160±0.0999	0.0805±0.0146
	0.05	0.0786±0.0102	0.0775±0.0112	0.0996±0.0089	0.0780±0.0104	0.0823±0.0083
	0.10	0.1086±0.0153	0.1360±0.0189	0.1570±0.0170	0.1218±0.0118	0.1379±0.0169
	0.20	0.1395±0.0233	0.2233±0.0253	0.2462±0.0285	0.1900±0.0227	0.2195±0.0328
WFG5	0.00	1.1024±0.1760	0.0690±0.0002	0.0747±0.0023	1.1440±0.2605	0.3017±0.0503
	0.05	0.0950±0.0088	0.0910±0.0062	0.0976±0.0052	0.0905±0.0039	0.0913±0.0042
	0.10	0.1294±0.1202	0.1163±0.0098	0.1271±0.0118	0.1077±0.0085	0.1225±0.0119
	0.20	0.1290±0.0440	0.1700±0.0204	0.1951±0.0315	0.1414±0.0125	0.1737±0.0207
WFG6	0.00	0.8822±0.4627	0.1718±0.0776	0.1447±0.0808	1.0289±0.3312	0.3404±0.0275
	0.05	0.1461±0.0456	0.1996±0.0484	0.1788±0.0593	0.2274±0.0410	0.2201±0.0418
	0.10	0.1876±0.0482	0.2468±0.0242	0.2407±0.0324	0.2518±0.0116	0.2496±0.0088
	0.20	0.2709±0.1098	0.2971±0.0283	0.3058±0.0217	0.2759±0.0177	0.3031±0.0247
WFG7	0.00	0.7434±0.1007	0.4237±0.1247	0.2981±0.0840	1.0457±0.1247	0.6685±0.1277
	0.05	0.9066±0.5840	0.7143±0.1499	0.3694±0.1349	1.2833±0.6606	0.9653±0.5107
	0.10	1.5865±0.5823	1.7014±0.4998	0.9249±0.6724	1.9126±0.2070	1.6610±0.5213
	0.20	1.8447±0.4069	1.8096±0.3532	1.6331±0.5247	1.9120±0.0401	1.9152±0.1518
WFG8	0.00	1.4918±0.3192	0.5918±0.0315	0.6414±0.0184	1.4618±0.3574	0.9085±0.1867
	0.05	0.6682±0.0334	0.7179±0.0248	0.7078±0.0217	0.7167±0.0259	0.7264±0.0250
	0.10	0.7485±0.0463	0.8066±0.0359	0.7988±0.0226	0.8128±0.0401	0.8142±0.0286
	0.20	0.8360±0.0372	0.9315±0.0296	0.8837±0.0215	0.9408±0.0273	0.9521±0.0293
WFG9	0.00	0.5196±0.1438	0.2392±0.0021	0.2078±0.0654	0.6068±0.2152	0.3989±0.0508
	0.05	0.2547±0.0211	0.2519±0.0054	0.2432±0.0344	0.2520±0.0054	0.2533±0.0077
	0.10	0.2520±0.0062	0.2674±0.0094	0.2717±0.0153	0.2646±0.0113	0.2722±0.0116
	0.20	0.2568±0.0116	0.3098±0.0157	0.3100±0.0222	0.2873±0.0140	0.3082±0.0272

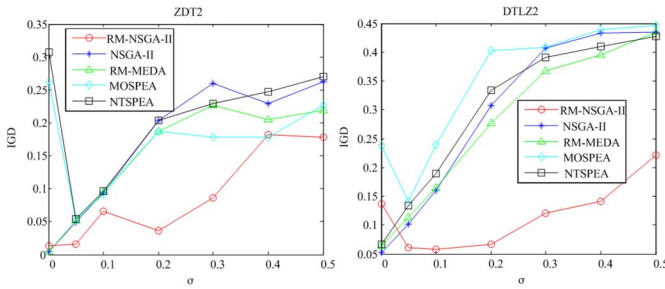


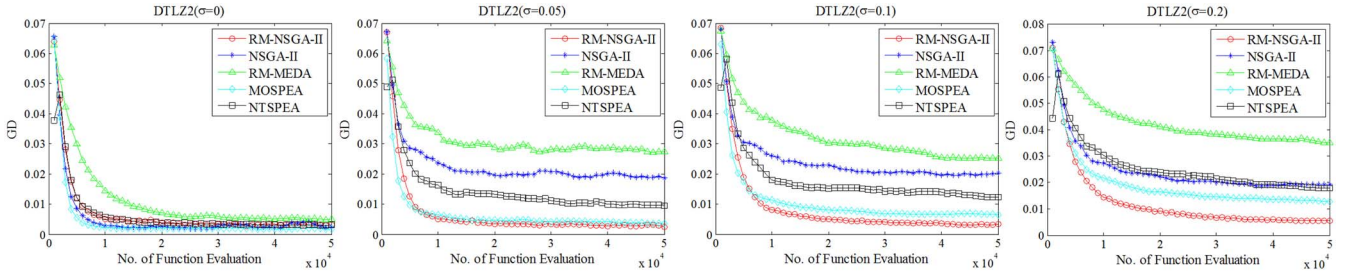
Fig. 11. IGD values of RM-NSGA-II, NSGA-II, RM-MEDA, MOSPEA, and NTSPEA on ZDT2 and DTLZ2 in different noise levels (up to $\sigma = 0.5$).

IGD values on noisy problems. For the two-objective problem ZDT2 with noise $\sigma = 0.2$, algorithms expect for RM-NSGA-II have similar IGD values. When σ grows to 0.4, MOEAs for noisy problems (MOSPEA and NTSPEA) perform better than MOEAs for noise-free problems (NSGA-II and RM-MEDA), as expected. However, RM-NSGA-II has a significant advantage over all those four algorithms. Interestingly, after $\sigma = 0.4$, NTSPEA cannot beat NSGA-II, and the advantage of RM-NSGA-II is not as significant as the case of a smaller σ . For the 3-objective problem DTLZ2, RM-NSGA-II increases its IGD values much slower than the other four algorithms. In short,

RM-NSGA-II can handle problems with noises, large or small, better than all other algorithms.

2) *Discussion:* From the results in the last section, we can conclude the behaviors of the characteristics on these compared algorithms. Fig. 12 is the GD values of RM-NSGA-II, NSGA-II, RM-MEDA, MOSPEA, and NTSPEA over generations on DTLZ2 ($\sigma = 0$, $\sigma = 0.05$, $\sigma = 0.1$, and $\sigma = 0.2$). For the noise-free DTLZ2, all the compared algorithms have approximated convergence performance except for RM-MEDA. However, the situation changes when σ increases. In the cases with $\sigma = 0.05$ and $\sigma = 0.1$, MOEAs for noisy MOPs (RM-NSGA-II, MOSPEA, and NTSPEA) have better convergence performance than MOEAs for noisy-free MOPs (NSGA-II and RM-MEDA), MOSPEA converges fast in the first 10 000 function evaluations, but RM-NSGA-II outperforms MOSPEA after 10 000 function evaluations; NTSPEA has a smaller GD value than NSGA-II but a larger GD value than RM-NSGA-II and MOSPEA. When σ grows to 0.2, the advantage of RM-NSGA-II over MOSPEA on GD rises, and NTSPEA cannot outperform NSGA-II due to the large number of re-evaluations for the high level of noises.

NSGA-II searches individuals to form the nondominated solution set, while RM-NSGA-II obtains a regularity model of the nondominated solution set. Therefore, NSGA-II can

Fig. 12. GD values of RM-NSGA-II, NSGA-II, RM-MEDA, MOSPEA, and NTSPEA over generations on DTLZ2 ($\sigma = 0$, $\sigma = 0.05$, $\sigma = 0.1$, and $\sigma = 0.2$).TABLE IX
COMPARISON BETWEEN RM-NSGA-II AND RM-MEDA

	RM-NSGA-II	RM-MEDA
Set to Build Models	Non-dominated Set	Whole Population
Model	ζ	$\zeta + \varepsilon$
Cluster Number	Self-Adaptive	Fixed in Advance
Sampling Extent	$[a_i, b_i]$	$[a_i - 0.25(b_i - a_i), b_i + 0.25(b_i - a_i)]$
Endpoint Maintenance	Applied	None
Diversity Maintenance	Uniform Sampling Matrix	Random Sampling Matrix
De-noising	Extra De-noising	None

obtain a small number of solutions very close to the true PFs but distributed randomly. As NSGA-II has no de-noising strategy, it cannot obtain the observed objective values precisely or maintain diversity reasonably, which is the reason why the performance of NSGA-II on noisy problems is unstable. In contrast, RM-NSGA-II considers the solution set as a whole, which helps de-noising and obtaining a solution set with both satisfactory convergence and diversity. However, building $(m - 1)$ -D models in the decision space may limit RM-NSGA-II. There is a mapping relation from PS to PF, a large spread of PS might not lead to a large spread of PF, which is the reason why RM-NSGA-II cannot have a larger spread of PF than NSGA-II on some WFG problems (their solution sets are more complicated than that of ZDT problems, and NSGA-II emphasizes extreme points in the objective space), WFG2 for instance. That is also the reason why RM-NSGA-II loses its performance on the multimodal problem DTLZ3.

RM-MEDA uses the regularity model as RM-NSGA-II, but it is not suitable for noisy MOPs from the result. From Table IX, we can find the significant differences between RM-NSGA-II and RM-MEDA.

For noisy MOPs, RM-NSGA-II only uses nondominated solutions rather than the whole population as RM-MEDA does to build the model, because RM-NSGA-II aims to capture the manifold of the nondominated set. In contrast, RM-MEDA aims to learn the manifold of the population by the regularity model to promote optimization. For such different aims, RM-MEDA introduces noise ε in the model and extends the manifold to add diversity, whereas RM-NSGA-II only keeps the manifold strictly from the obtained model. RM-NSGA-II also adds special strategies to maintain endpoints and to de-noise, which RM-MEDA never takes into account. Additionally, RM-NSGA-II identifies the number of clusters adaptively and maintains the sampling matrix into an USM. All the above differences make RM-NSGA-II more effective than RM-MEDA in solving noisy MOPs.

For noise-free problems, the ED strategy in RM-NSGA-II still averages the objective values without uncertainty of different solutions in a small neighborhood, which can lead to inaccurate objective values for RM-NSGA-II, thus the performance of RM-NSGA-II is lowered (shown in Section VI-B). That is the reason why RM-NSGA-II cannot perform better than RM-MEDA on the problems without any noises.

MOSPEA is a representative of MOEAs for noisy MOPs with the probabilistic Pareto ranking, which results in its good convergence ability on noisy problems. That is the reason why MOSPEA performs better on hard problems such as DTLZ1, DTLZ3, and ZDT4. However, probabilistic Pareto ranking cannot provide satisfactory diversity in MOSPEA, especially when the size of noises increases. Both RM-NSGA-II and MOSPEA aim at noisy MOPs, hence we only compare their performance on noisy problems. Probabilistic Pareto ranking performs better than building a model on convergence when the structure of PS is simple (see the results on the ZDT and DTLZ problems). However, when the PS becomes complicated, building a model works better than probabilistic Pareto ranking.

NTSPEA adopts re-evaluation, thus, the effect brought by noises can be reduced. With relatively accurate objective evaluations, NTSPEA can obtain the solutions with better convergence even on DTLZ3, whereas RM-NSGA-II cannot obtain approximated models. However, the model construction in RM-NSGA-II works on the WFG problems, which leads to better convergence than NTSPEA.

Summarily, existing MOEAs such as NSGA-II and RM-MEDA are insufficient to solve noisy MOPs because they were not designed to cope with noises in MOPs. For the two compared MOEAs for noisy MOPs, both of them fail to maintain diversity well due to their focusing on a single individual rather than a model of a solution set. Comparing with these algorithms, RM-NSGA-II improves NSGA-II on noisy MOPs. The regularity model can efficiently learn from

noisy environments, but its performance on the spread may be limited by the $(m - 1)$ -D model. Therefore, it is not good at multimodal problems.

VII. CONCLUSION

In this paper, we have analyzed the de-noising performance of the regularity model in noisy environments and the behavior of the regularity model embedded in NSGA-II (RM-NSGA-II) for noisy MOPs. Due to the effectiveness of the regularity model, RM-NSGA-II can obtain the solution set with both satisfactory convergence and diversity in noisy environments, which is shown by our experiments. The contributions of this paper are summarized as follows.

- 1) *De-Noising Ability of Regularity Model*: Although the regularity model has been applied in RM-MEDA, it is the first time used for handling noisy MOPs. In this paper, we find the effectiveness of the regularity model in noisy MOPs. The reason comes from two aspects, one is from the natural de-noising characteristics of modeling, the other one is from the dimension-reduced complexity by considering the features of MOPs.
- 2) *Improving Existing MOEAs on Noisy Problems*: The regularity model has very good transportability in existing MOEAs. Thus, those MOEAs that are not good at noisy problems can be improved by embedding the regularity model. With the regularity model like a patch in the system of MOEAs, existing MOEAs can solve noisy MOPs.

Although the regularity model can help existing MOEAs to solve noisy MOPs satisfactorily, there are still several issues that should be studied in the future.

- 1) The convergence of RM-NSGA-II on multimodal problems such as DTLZ3 is not good enough.
- 2) In the regularity model, only the even disturbance around PS is assumed. Hence, the regularity model may not perform well for discontinuous MOPs or MOPs with unsymmetrical mapping from the decision space to the objective space.
- 3) The diversity exploration ability of the regularity model on the WFG problems should be improved.

REFERENCES

- [1] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.
- [2] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Trans. Evol. Comput.*, vol. 11, no. 6, pp. 712–731, Dec. 2007.
- [3] Q. Zhang, A. Zhou, and Y. Jin, "RM-MEDA: A regularity model-based multiobjective estimation of distribution algorithm," *IEEE Trans. Evol. Comput.*, vol. 12, no. 1, pp. 41–63, Feb. 2008.
- [4] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm," in *Proc. Evol. Methods Design Optim. Control Appl. Ind. Probl. (EUROGEN)*, Athens, Greece, 2001, pp. 1–21.
- [5] K. Li, S. Kwong, Q. Zhang, and K. Deb, "Interrelationship-based selection for decomposition multiobjective optimization," *IEEE Trans. Cybern.*, to be published.
- [6] L. Ke, Q. Zhang, and R. Battiti, "MOEA/D-ACO: A multiobjective evolutionary algorithm using decomposition and AntColony," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1845–1859, Dec. 2013.
- [7] X. Ma *et al.*, "A multiobjective evolutionary algorithm based on decision variable analyses for multi-objective optimization problems with large scale variables," *IEEE Trans. Evol. Comput.*, to be published.
- [8] P. Boonma and J. Suzuki, "A confidence-based dominance operator in evolutionary algorithms for noisy multiobjective optimization problems," in *Proc. 21st Int. Conf. Tools AI (ICTAI)*, Newark, NJ, USA, 2009, pp. 387–394.
- [9] H. Eskandari and C. D. Geiger, "Evolutionary multiobjective optimization in noisy problem environments," *J. Heuristics*, vol. 15, no. 6, pp. 559–595, Dec. 2009.
- [10] C. Goh and K. Tan, "Noise handling in evolutionary multi-objective optimization," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Vancouver, BC, Canada, 2006, pp. 1354–1361.
- [11] E. J. Hughes, "Evolutionary multi-objective ranking with uncertainty and noise," in *Evolutionary Multi-Criterion Optimization*. Berlin, Germany: Springer, 2001, pp. 329–343.
- [12] M. Babbar, A. Lakshmikantha, and D. E. Goldberg, "A modified NSGA-II to solve noisy multiobjective problems," in *Proc. Genet. Evol. Comput. Conf. (GECCO)*, Chicago, IL, USA, 2003, pp. 21–27.
- [13] M. Basseur and E. Zitzler, "A preliminary study on handling uncertainty in indicator-based multiobjective optimization," in *Applications of Evolutionary Computing*. Berlin, Germany: Springer, 2006, pp. 727–739.
- [14] D. Buche, P. Stoll, R. Dornberger, and P. Koumoutsakos, "Multiobjective evolutionary algorithm for the optimization of noisy combustion processes," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 32, no. 4, pp. 460–473, Nov. 2002.
- [15] A. Syberfeldt, A. Ng, R. I. John, and P. Moore, "Evolutionary optimisation of noisy multi-objective problems using confidence-based dynamic resampling," *Eur. J. Oper. Res.*, vol. 204, no. 3, pp. 533–544, Aug. 2010.
- [16] A. Singh, "Uncertainty based multi-objective optimization of ground-water remediation design," Ph.D. dissertation, Univ. Illinois Urbana-Champaign, Urbana, IL, USA, 2003.
- [17] H. Tang, V. A. Shim, K. C. Tan, and J. Y. Chia, "Restricted Boltzmann machine based algorithm for multi-objective optimization," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Barcelona, Spain, 2010, pp. 1–8.
- [18] V. A. Shim, K. C. Tan, J. Y. Chia, and A. Al Mamun, "Multi-objective optimization with estimation of distribution algorithm in a noisy environment," *Evol. Comput.*, vol. 21, no. 1, pp. 149–177, 2013.
- [19] Y. Hong, Q. Ren, and J. Zeng, "Optimization of noisy fitness functions with univariate marginal distribution algorithm," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, vol. 2, Edinburgh, U.K., 2005, pp. 1410–1417.
- [20] K. Miettinen, *Nonlinear Multiobjective Optimization*, vol. 12. New York, NY, USA: Springer, 1999.
- [21] O. Schütze, S. Mostaghim, M. Dellnitz, and J. Teich, "Covering Pareto sets by multilevel evolutionary subdivision techniques," in *Evolutionary Multi-Criterion Optimization*. Berlin, Germany: Springer, 2003, pp. 118–132.
- [22] N. Kambhatla and T. K. Leen, "Dimension reduction by local principal component analysis," *Neural Comput.*, vol. 9, no. 7, pp. 1493–1516, Oct. 1997.
- [23] K. Deb and S. Tiwari, "Omni-optimizer: A procedure for single and multi-objective optimization," in *Evolutionary Multi-Criterion Optimization*. Berlin, Germany: Springer, 2005, pp. 47–61.
- [24] K. McClymont and E. Keedwell, "Deductive sort and climbing sort: New methods for non-dominated sorting," *Evol. Comput.*, vol. 20, no. 1, pp. 1–26, 2012.
- [25] H. Wang and X. Yao, "Corner sort for Pareto-based many-objective optimization," *IEEE Trans. Cybern.*, vol. 44, no. 1, pp. 92–102, Jan. 2014.
- [26] X. Zhang, Y. Tian, R. Cheng, and Y. Jin, "An efficient approach to non-dominated sorting for evolutionary multiobjective optimization," *IEEE Trans. Evol. Comput.*, vol. 19, no. 2, pp. 201–213, Apr. 2015.
- [27] R. B. Agrawal, K. Deb, and R. B. Agrawal, "Simulated binary crossover for continuous search space," *Complex Syst.*, vol. 9, no. 2, pp. 115–148, 1994.
- [28] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin, "The elements of statistical learning: Data mining, inference and prediction," *Math. Intell.*, vol. 27, no. 2, pp. 83–85, 2005.
- [29] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, vol. 4. Englewood Cliffs, NJ, USA: Prentice-Hall, 1992.
- [30] Z. Wang, K. Tang, and X. Yao, "Multi-objective approaches to optimal testing resource allocation in modular software systems," *IEEE Trans. Rel.*, vol. 59, no. 3, pp. 563–575, Sep. 2010.
- [31] H. Wang, L. Jiao, R. Shang, S. He, and F. Liu, "A memetic optimization strategy based on dimension reduction in decision space," *Evol. Comput.*, vol. 23, no. 1, pp. 69–100, 2015.

- [32] E. Zitzler, K. Deb, and L. Thiele, "Comparison of multiobjective evolutionary algorithms: Empirical results," *Evol. Comput.*, vol. 8, no. 2, pp. 173–195, Jun. 2000.
- [33] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable multi-objective optimization test problems," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Honolulu, HI, USA, 2002, pp. 825–830.
- [34] S. Huband, P. Hingston, L. Barone, and L. While, "A review of multiobjective test problems and a scalable test problem toolkit," *IEEE Trans. Evol. Comput.*, vol. 10, no. 5, pp. 477–506, Oct. 2006.
- [35] D. A. Van Veldhuizen and G. B. Lamont, "On measuring multiobjective evolutionary algorithm performance," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, vol. 1. La Jolla, CA, USA, 2000, pp. 204–211.
- [36] S. Bandyopadhyay, S. K. Pal, and B. Aruna, "Multiobjective GAs, quantitative indices, and pattern classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 5, pp. 2088–2099, Oct. 2004.
- [37] Q. Zhang *et al.*, "Multiobjective optimization test instances for the CEC 2009 special session and competition," School Comput. Sci. Electron. Eng., Univ. Essex, Colchester, U.K., Dept. Mech. Eng., Nanyang Technol. Univ., Singapore, Tech. Rep. CES-487, Apr. 2009, pp. 1–30.
- [38] M. Hollander and D. Wolfe, *Nonparametric Statistical Methods*. New York, NY, USA: Wiley, 1999.



Handing Wang (S'10) received the B.Eng. in intelligence science and technology and Ph.D. degree in circuit and system from Xidian University, Xi'an, China, in 2010 and 2015, respectively.

She is currently a Research Fellow with the Department of Computing, University of Surrey, Guildford, U.K. Her current research interests include nature-inspired computation, multiobjective optimization, multiple criteria decision making, and real-world problems.



Qingfu Zhang (M'01–SM'06) received the B.Sc. degree in mathematics from Shanxi University, Taiyuan, China, in 1984, and the M.Sc. degree in applied mathematics and the Ph.D. degree in information engineering from Xidian University, Xi'an, China, in 1991 and 1994, respectively.

He is a Professor with the Department of Computer Science, City University of Hong Kong, Hong Kong, and the School of Computer Science and Electronic Engineering, University of Essex, Colchester, U.K., and a Changjiang Visiting Chair

Professor with Xidian University. He is currently leading the Metaheuristic Optimization Research Group, City University of Hong Kong. His current research interests include evolutionary computation, optimization, neural networks, data analysis, and their applications. He holds two patents and has authored several research publications.

Dr. Zhang was a recipient of the Unconstrained Multiobjective Optimization Algorithm Competition at the Congress of Evolutionary Computation 2009 for a multiobjective optimization algorithm developed in his group and the 2010 IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION Outstanding Paper Award. He is an Associate Editor of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION and the IEEE TRANSACTIONS ON CYBERNETICS. He is an Editorial Board Member of three other international journals.



Licheng Jiao (SM'89) received the B.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 1982, and the M.S. and Ph.D. degrees in electronic engineering from Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively.

Since 1992, he has been a Professor with the School of Electronic Engineering, Xidian University, Xi'an, where he is currently the Director of the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China. He is in charge of about 40 important scientific research projects and has published over 20 monographs and 100 papers in international journals and conferences. His current research interests include image processing, natural computation, machine learning, and intelligent information processing.

Dr. Jiao is the President of the Computational Intelligence Chapter, the IEEE Xi'an Section, and the IET Xi'an Network, the Chairman of the Awards and Recognition Committee, the Vice Board Chairperson of the Chinese Association of Artificial Intelligence, a Councilor of the Chinese Institute of Electronics, a Committee Member of the Chinese Committee of Neural Networks, and an expert of the Academic Degrees Committee of the State Council.

Dr. Jiao is the President of the Computational Intelligence Chapter, the IEEE Xi'an Section, and the IET Xi'an Network, the Chairman of the Awards and Recognition Committee, the Vice Board Chairperson of the Chinese Association of Artificial Intelligence, a Councilor of the Chinese Institute of Electronics, a Committee Member of the Chinese Committee of Neural Networks, and an expert of the Academic Degrees Committee of the State Council.



Xin Yao (M'91–SM'96–F'03) received the B.Sc. degree from the University of Science and Technology of China (USTC), Hefei, China, in 1982, the M.Sc. degree from the North China Institute of Computing Technology, Beijing, China, in 1985, and the Ph.D. degree in computer science from the USTC, in 1990.

He was an Associate Lecturer and a Lecturer with the USTC from 1985 to 1990, a Post-Doctoral Fellow with Australian National University, Canberra, ACT, Australia, and Commonwealth

Scientific and Industrial Research Organization, Melbourne, VIC, Australia, from 1990 to 1992, and a Lecturer, a Senior Lecturer, and an Associate Professor with the University of New South Wales, Sydney, NSW, Australia, and the Australian Defence Force Academy, Canberra, from 1992 to 1999. Since 1999, he has been a Professor (Chair) of Computer Science with the University of Birmingham, Birmingham, U.K., where he is currently the Director of the Centre of Excellence for Research in Computational Intelligence and Applications.

Mr. Yao was a recipient of the 2001 IEEE Donald G. Fink Prize Paper Award, the 2010 IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION Outstanding Paper Award, the 2011 IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award, and several other best paper awards. He is an IEEE Computational Intelligence Society Distinguished Lecturer and a former Editor-in-Chief of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION.